

H2O.ai



H2O
**FEATURE
STORE**

Co-developed with



AT&T

PRODUCT SHEET



Co-developed with

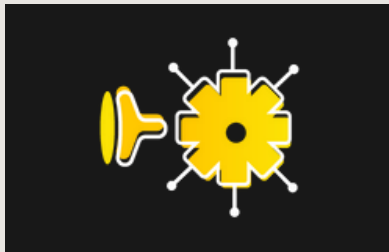


H2O.ai

KEY FEATURES

Companies around the world are driven to grow revenue, optimize operations, mitigate risk, and personalize experiences. In this modern era, the data needed to understand trends and predict outcomes are captured and stored digitally across a variety of tools and storage repositories making it difficult to gain a comprehensive, well-informed view by which to make decisions.

The H2O Feature Store intelligently extracts, manages and optimizes the feature engineering and model building process by seamlessly connecting information across platforms, establishing quality, consistency and transparency across the machine learning lifecycle and ultimately bringing together people with the diversity of skill sets needed to deliver transformational value with AI.



Seamless data ingestion

- Connect any engineering pipeline via a REST API.
- Out-of-the-box integrations available for Snowflake, Databricks, H2O Sparkling Water, Apache Spark, and more.



Automated feature transformation

- Automated feature engineering, encoding and selection.
- Per-feature controls.
- Automated validation and cross-validation of features.



Intelligent registry management

- Detailed cataloging of over 40 metadata attributes, such as description, data sources, and data sensitivity categories.
- Explore and find features based on automatically generated feature statistics like mean, standard deviation, frequencies etc.



Recommendations & rankings

- New features and feature updates are automatically generated and recommended to improve model performance.
- Recommendations run automatically or on-demand.
- Scores indicate popularity or value across different use-cases.
- Statistics like mean, median, standard deviation, frequencies etc. provide quick insights on features.



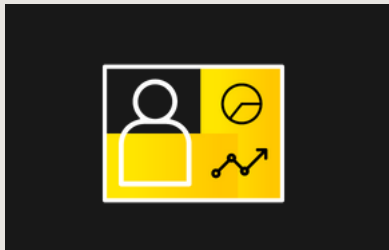
Semantic search and discovery

- Unsupervised search of features available out-of-the-box.
- Active learning from added labels and ontologies improves results.
- Feedback mechanism in the UI adds for continuous improvement.
- Discover related features based on search parameters.



Flexible feature serving

- Batch
- In-stream (submillisecond latency)
- On demand



Persistent model monitoring

- Analyze and report on detected bias in features.
- Automatically checks both individual features and feature sets for drift over time and alerts anomalous behavior to users.
- Alerts can be used to trigger retraining or refitting to keep models accurate.



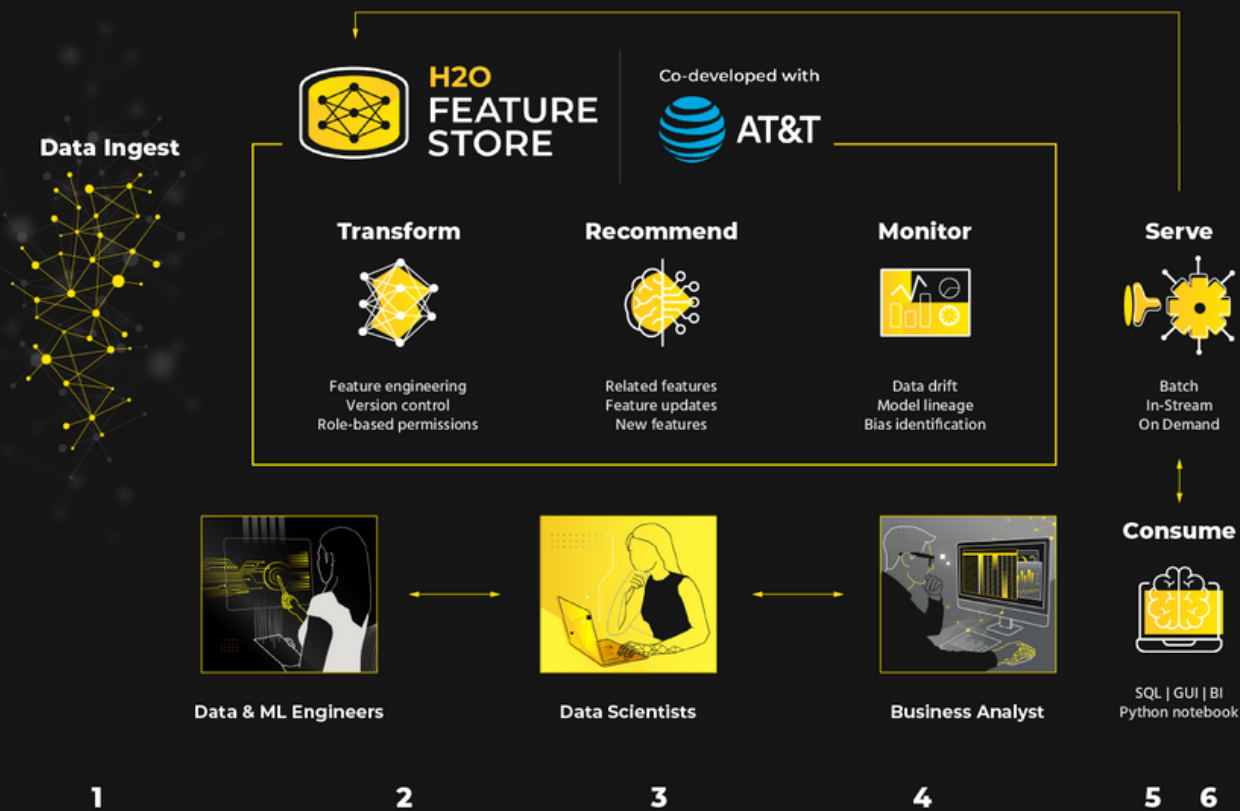
Access management and governance control

- Integrate existing identity and access management tools.
- Grant role-based and individual permissions to specific projects and features.
- Store versions of features and their associated metadata to comply with regulations, backtest models, and learn from past features.
- Automatically log changes to all feature values to understand how past models would have scored records at any given time.



Flexible integrations & deployment options

- Both notebook and UI interfaces.
- Integrated with Snowflake, Databricks, H2O Sparkling Water, Apache Spark, Python, Java, and Scala feature engineering pipelines.
- REST API access.
- Seamless H2O AI Cloud integration.
- Cloud agnostic.
- Features stored in-memory and served with sub-millisecond latency.
- Kubernetes-based deployment to support scaling needs.



1

Data Ingest

Data pipelines build a connection between the information stored in various data repositories to a single feature store.

2

Transform

Data scientists then optimize the data with feature engineering, which is used to extract the most signals from a given dataset to drive machine learning models. Centralized access and control across the feature engineering and model building process means quality and consistency are tracked and maintained across feature sets.

3

Recommend

The feature store uses a smart matching engine to proactively recommend high quality, already engineered features to use based on the unique data and business case a user is working on.

4

Monitor

Machine learning algorithms then actively monitor and provide alerts about data quality and drift to ensure users are aware when model performance might be impacted in any way.

5

Serve

Data can be accessed in batch, in stream, or on-demand as required. Data and machine learning engineers can also configure H2O Feature Store to perform data quality tasks and maintain version control over those tasks, to ensure data scientists and business users alike are accessing the best quality version of the data needed to drive value.

6

Consume

Business users can gain insight into which pieces of information are influencing predictions, the importance of specific features in driving those predictions, and, using subject matter expertise, they can adjust the information being used and how it is being interpreted by the models surfacing those insights.